A/B Testing

Author: Palak Sarawagi

Differences in differences regression (DID)

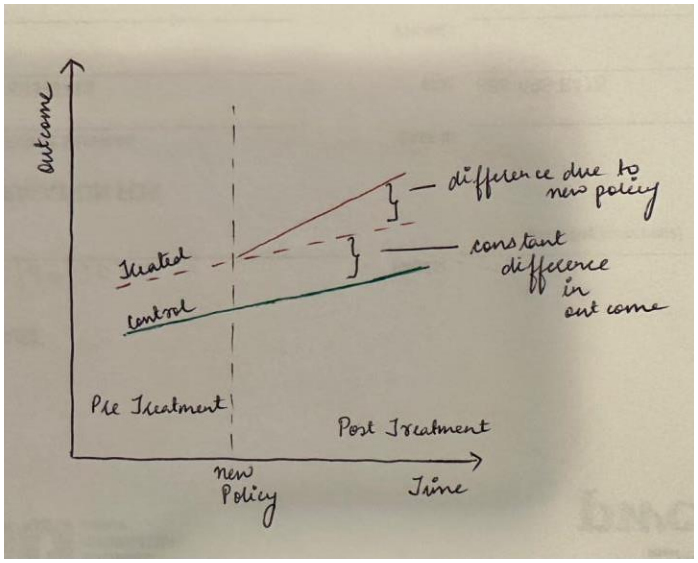
Differences in differences regression is used to calculate the effect of a treatment by comparing the outcome of both the control group and a treated group. For DID, we also consider the situation if the treatment was not done: what would be the outcome in control and treated group?

DID regression is usually used to study the causal effect of a treatment in a natural experiment where the treated and controlled groups are not determined randomly.

For example, we can use DID regression to determine the effect of a new school policy where teaching assistance is provided to some children. We can evaluate the effect by comparing the grades for students for both the treated and controlled group before and after the new policy.

DID take the assumption that the treated and controlled group follow the same trend (parallel) before the treatment. So, the change in outcome for both the treated and controlled group can be attributed to the treatment itself.

Let us understand the same using this diagram below:



Mathematically, it can be calculated as

DID= (Total(treated)|after- Total(control)|after) – (Total(treated)|before-Total(control)|before)

The DID model is represented below as a simple OLS regression:

$Total_{i,t} = \alpha + \beta * after_t + \gamma * treated_i + \delta * after_t * treated_i + \varepsilon_{it}$

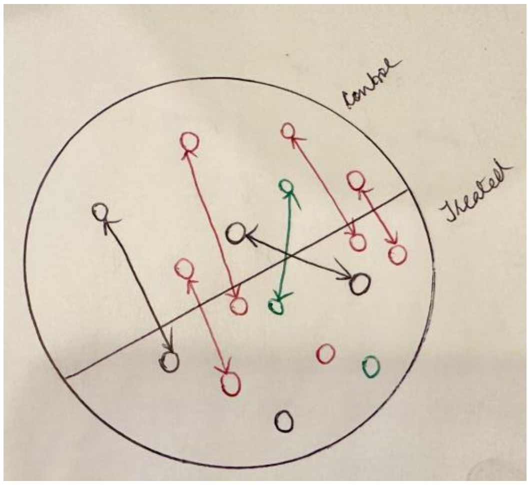
The coefficient δ tells the causal effect.

Propensity Score Matching

As mentioned above, the treated and controlled groups are not determined randomly in a natural experiment. So, the groups can be biased in the above regression due to other covariants. Propensity Score Matching is used to remove such bias by comparing the units in the treated and controlled group. It aims to control the bias by making the treatment and control group comparable with respect to other covariants.

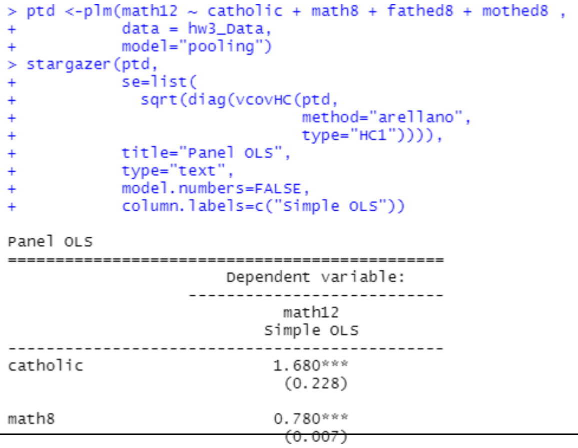
In other words, we create an artificial control group by matching treated entities to control entities with similar characteristics. The treated entities which do not match with any control entities are not considered in the experiment. So in a way, PSM tries to achieve randomization artificially.

For example, in the below diagram, we can see that out of 10 units in the treated group, we found 7 control units similar to the treated units. However, the remaining 3 units in the treated group which has no similar entity in the control group. These 3 treated entities are not considered in our experiment when we run a regression to determine the causal effect.



Example:

Using a simple OLS regression model to find the effect of attending a catholic high school` on the math score in the 12th grade (Showing only the relevant O/P in this snapshot)



```
> ptd <-plm(math12 ~ catholic + math8 + fathed8 + mothed8 ,
+           data = hw3_Data,
+           model="pooling")
> stargazer(ptd,
+           se=list(
+             sqrt(diag(vcovHC(ptd,
+                              method="arellano",
+                              type="HC1")))),
+           title="Panel OLS",
+           type="text",
+           model.numbers=FALSE,
+           column.labels=c("Simple OLS"))

Panel OLS
==============================================
                      Dependent variable:
                  ----------------------------
                             math12
                           Simple OLS
----------------------------------------------
catholic                    1.680***
                             (0.228)

math8                       0.780***
                             (0.007)
```

It can be determined from the above model that with a per unit increase in the catholic school variable, the match12 score increases by 1.680 units.

However, the OLS result may hardly be interpreted as a causal effect. It reasons include but are not limited to the following:

(i) The treated and controlled group are not randomized. Therefore, the relationship could be related to other covariants like additional home tutoring or better access to Wi-Fi and books. But since we do not have data for these variants, it is part of the error term. When the independent variable is correlated to the error term, its coefficient in the regression does not produce the desired causal effect.
(ii) The model does not consider the unobserved time-variant coefficients.
(iii) The model has not considered the effect of variables such as race into consideration.
(iv) The data in the treated and control group is not equal so there is a selection bias in the data too.

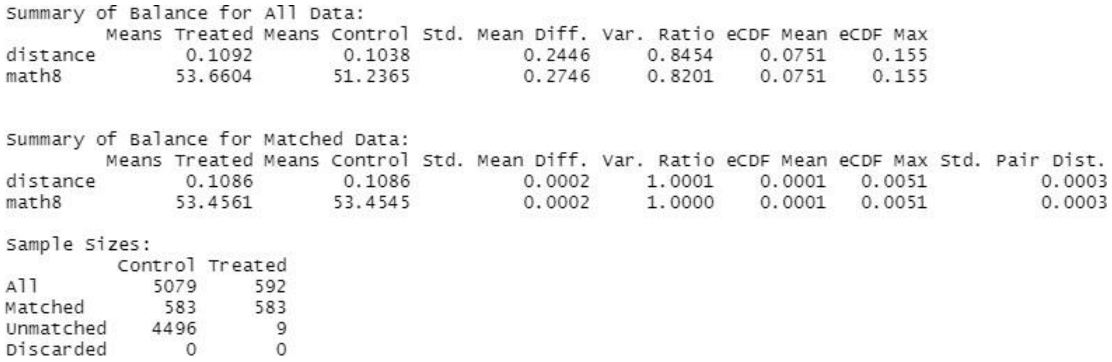
Using Propensity Score Matching (PSM) to find the effect of attending a catholic high school onthe math score in the 12th grade.

The following code is used to do the Propensity score matching.

```
> matched_dataset <- matchit(catholic ~ math8, data = hw3_Data, method = 'neaíest', calipeí = 0.003)
```

```
> summaíy(matched_dataset)
```

Output:



```
Summary of Balance for All Data:
         Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max
distance        0.1092        0.1038          0.2446     0.8454    0.0751    0.155
math8          53.6604       51.2365          0.2746     0.8201    0.0751    0.155


Summary of Balance for Matched Data:
         Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max Std. Pair Dist.
distance        0.1086        0.1086          0.0002     1.0001    0.0001   0.0051          0.0003
math8          53.4561       53.4545          0.0002     1.0000    0.0001   0.0051          0.0003

Sample Sizes:
          Control Treated
All          5079     592
Matched       583     583
Unmatched    4496       9
Discarded       0       0
```

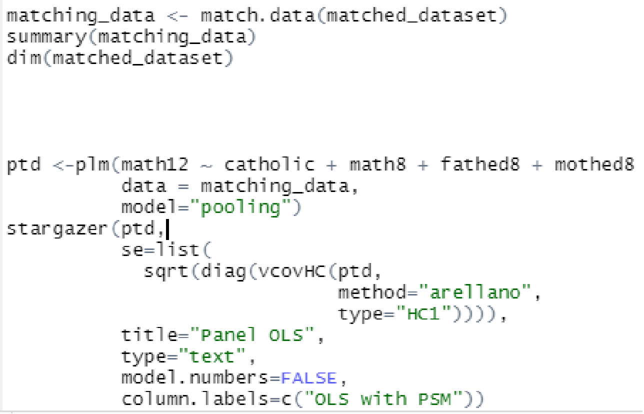
After doing a PSM using the nearest neighbor algorithm with a caliper of 0.003, only 583 of the 592 treated units were matched with control groups.

In doing the experiment further, we will only include the treated units with a matched controlled unit. Therefore, we removed 4496 control and 9 treated data units from our further experiment. We now have balanced data to run the OLS regression model.

It is to note here that the mean of math8 scores for both the treated and control group is nearly equal, so we can see higher internal validity. However, it also reduced the external validity because we can no longer make claims for the dropped 4496 units of data.

Let us now run the OLS regression model on the matched data (post PSM).

The following code is used to extract the data and run the model:

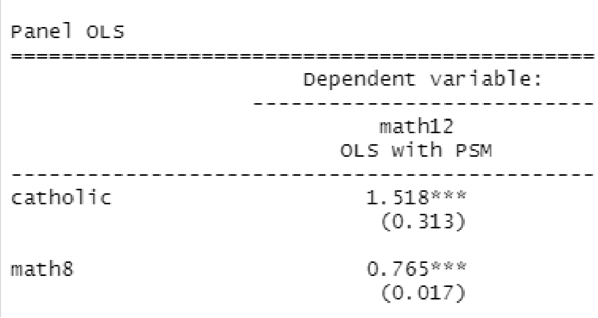


```
matching_data <- match.data(matched_dataset)
summary(matching_data)
dim(matched_dataset)




ptd <-plm(math12 ~ catholic + math8 + fathed8 + mothed8
          data = matching_data,
          model="pooling")
stargazer(ptd,
          se=list(
            sqrt(diag(vcovHC(ptd,
                             method="arellano",
                             type="HC1")))),
          title="Panel OLS",
          type="text",
          model.numbers=FALSE,
          column.labels=c("OLS with PSM"))
```

The following output was observed:



```
Panel OLS
===============================================
                         Dependent variable:
                     ---------------------------
                               math12
                            OLS with PSM
-----------------------------------------------
catholic                      1.518***
                              (0.313)

math8                         0.765***
                              (0.017)
```

It can be determined from the above model that with a per unit increase in the catholic school variable, the match12 score increases by 1.518 units.

This model is a better estimation of a causal effect than the simple regression model because it eliminates the bias by making the treatment and control group comparable with respect to other covariants.